# Assessment Strategies for Pair Programming

***Jan Hendrik Hahn, Elsa Mentz, and Lukas Meyer***
***North-West University, Potchefstroom Campus,***
***Potchefstroom, South Africa***

**Hendrik.Hahn@nwu.ac.za; Elsa.Mentz@nwu.ac.za;**
**Lukas.Meyer@nwu.ac.za**

## Executive Summary

Although pair programming has proved its usefulness in teaching and learning programming skills, it is difficult to assess the individual roles and abilities of students whilst programming in pairs. (Note that within this manuscript, the term assessment refers to evaluating individual student performance.)

Assessing only the outcomes of a pair programming assignment and awarding each member of the pair the same score may not be a reliable reflection of individual programming abilities; it may result in a discrepancy between students' individual and pair programming marks. The aim of our study was to answer the following question: Do combined self, peer, and facilitator assessment strategies for pair programming contribute towards a more reliable assessment of individual programming abilities? In our study, all three types of assessments were conducted after the completion of every pair programming assignment. In each case, we used rubrics containing specific assessment criteria. These rubrics were provided to students at the beginning of every pair programming assignment. The facilitator gave formative feedback to students after completion of a pair programming assignment and thereafter when an individual test was written.

To determine whether these specific assessment strategies could contribute towards a more reliable way of assessing individual abilities in pair programming situations, the averages of students' pair and individual marks were correlated. Moreover, two practical examinations of equal academic standard were conducted at the end of the course: one for pairs and the other for individuals. A significantly high correlation coefficient between the individual and pair examination marks was obtained, indicating that the pair assessment strategies implemented in this study proved to be successful in determining students' individual programming abilities.

The results emerging from this study support the implementation of specific assessment strategies to assess individual programming abilities during pair programming situations.

**Keywords:** pair programming, formative assessment, summative assessment, assessment strategies, individual assessment, pair assessment, self-assessment, peer assessment, facilitator assessment.

## Introduction

Pair programming has been common practice in the programming industry during the last three decades, but only recently did it start to draw the attention of facilitators as a teaching strategy. Pair



**Editor: Dion Goh**

programming implies that two programmers execute the programming task together. One of them 'drives' the development and the typing of code, while the other one 'navigates' or points out the programming direction for the driver and provides feedback on the programming code (Adams, Lubega, Walmsley & Williams, 2004).

Studies conducted on pair programming identified some of its advantages for the teaching-learning situation (Tomayko, 2002; Williams & Kessler, 2001; Williams & Upchurch, 2001). For example, it has been found that when programmers work in pairs, fewer errors are made than in individual programming situations (Tomayko, 2002), resulting in better programming performance, increased confidence, and decreased frustration levels of the programmers (VanDeGrift, 2004). A possible explanation for these findings could be that pair members help each other to solve the problem and complete the programming task together. Thus, there seems to be some agreement among researchers that pair programming could be a promising teaching strategy for teaching programming skills (VanDeGrift, 2004).

Despite the possible benefits of pair programming as a teaching strategy, some of its limitations for student assessment have been documented as well. One such frequently encountered limitation is that some students may receive undeserved credit for the successful completion of a program (McDowell, Hanks, & Werner, 2003). The assumption is that it is difficult to assess students' individual programming abilities reliably in pair programming situations. The first author of this article experienced a similar problem when implementing pair programming as a teaching strategy during a second year Information Technology (IT) course for student teachers. The students achieved high marks for their pair assignments, but significantly lower marks for their individual assignments. This begs the question whether the results of pair programming assignments are reliable indicators of students' individual programming abilities. Naude and Hörne (2006) even classify undeserved credit under the umbrella of 'cheating,' and, although cheating is a strong word, the researchers emphasize the seriousness of the situation.

Several researchers highlight the problem of assigning group marks to individual students (Parsons, 2004). It is necessary to determine the contribution of each individual member and, according to Parson (2004), this is not an easy task. Some students could contribute little or nothing at all, and if there is no assessment of students' individual contributions, marks awarded could be an inaccurate reflection of a student's abilities (Cheng & Warren, 2000). According to Parsons (2004) a fair mark allocated to a given student should reflect that individual's effort and abilities. This statement is supported by the research of Verhaart, Hagen, and Giles (2005), who also wished to determine whether students' marks in group assessments correlated with their marks in individual assessments. They proposed two different assessment methods as best practice for assessing an individual's performance in group work. In the *base mark adjusted* method, they give individual tests, as well as self- and peer assessments after every groupwork session. In the *task splitting method*, they split the groupwork task and require some work done individually and other work done in groups. A specific weighting is allocated to the different tasks and students need to include a peer review to adjust the group contribution. Verhaart et al. (2005) concluded that both these methods are "*valid assessment forms, producing marks which seem to reflect the students' typical level of achievement.*"

In view of the research being done on group assessment, the aim of this study was to determine whether the use of specific assessment strategies for pair programming could contribute towards a more reliable way of assessing individual abilities in pair programming situations. In the development of the assessment strategies, literature on assessment in general and on the assessment of co-operative and collaborative learning was considered. In order to determine whether these specific pair programming assessment strategies could be implemented more reliably for assessing individual programming abilities, students' pair and individual marks were compared throughout the course to determine whether a significant correlation existed between these marks. In the next section the literature that served as a basis for the specific assessment strategies will be discussed.

# Assessment Strategies

Assessment strategies that can be applied to group work will be discussed first, and afterwards guidelines will be considered for their specific application to pair programming.

Lambert and Lines (2000) define assessment as the process of collecting and recording, interpreting and applying information on students' performance in teaching and learning. Marneweck and Rouhani (2002), among others, distinguish between summative and formative assessment. Summative assessment takes place at the end of a learning experience, while formative assessment takes place during the course of the learning process. Black and Wiliam (1998) emphasize that assessment becomes formative when evidence is actually used to adapt teaching to meet students' needs. According to Luckett and Sutherland (2000) the aim of summative assessment is to determine students' achievement levels at the end of the learning experience. During the assessment of pair programming it is important to assess students' achievement levels at the completion of the program, but also during the pair programming activities in order to determine each member's contribution to the attainment of the outcomes (Williams & Kessler, 2003).

There are different methods of assessment (Ellis, 2001; Hanrahan & Isaacs, 2001; Marneweck & Rouhani, 2002). According to Marneweck and Rouhani (2002), **peer assessment** implies that learners are encouraged to help and assess each other during peer activities. Hanrahan and Isaacs (2001) describe peer assessment as a constructive contribution by members of a pair or group to assess their collaborative efforts. Verhaart et al. (2005) implemented peer assessment in both of their assessment models. Cheng and Warren (2000) also stressed the importance of peer assessment and proposed that students need to be actively involved in group assessments by determining their peers' contributions. Within a pair programming environment evidence also exists that peer assessment could be helpful in determining individual contributions. Williams, Wiebe, Yang, Ferzli, and Miller (2002) applied peer assessment on completion of every pair programming project, and Daniel (2003) stated that students never again complained about *a free ride* when implementing peer assessment in pair programming situations.

During **self-assessment** students assess their own work (Hanrahan & Isaacs, 2001). Self assessment could be valuable in encouraging students to take more responsibility for their own learning (Earl, 2003). Liebovich (2000) stated that students should use self-assessment to identify the skills they need to improve on. Individual accountability in group work, as well as in pair programming, plays a significant role (Johnson & Johnson, 2006; Preston, 2005) and self-assessment can encourage learners to take personal responsibility during group or pair activities. Williams and Upchurch (2001) also supported the view that students should assess their own contributions in the pair when performing pair programming.

**Individual assessment** takes place when learners complete an activity autonomously and are assessed individually by the facilitator (Knight, 2004). Johnson and Johnson (2006) strongly emphasised the role that individual assessment should play in the assessment of group work activities in order to eliminate 'free-riding.' Individual assessment implies that the facilitator assesses the activity and provides feedback to individual learners. Verhaart et al. (2005) proposed that individual assessments should contribute significantly towards students' marks during group work activities. One can therefore assume that individual assessments should take place after every pair programming session in order to assess individual performance more reliably.

**Facilitator assessment** is normally applied for assessing the quantity and accuracy of students' work, but could also take place during the completion of an assignment where a wide range of data could be collected to enable the facilitator to modify the learning process (Earl, 2003).

Different methods of assessment require different techniques. To implement self- and peer-assessment effectively, rubrics containing predetermined assessment criteria will help the facilitator and students to focus on the outcomes of the activity (Airasian, 2005). Rubrics can also help the facilitator give more productive feedback to learners. According to Earl (2003), feedback is part of the assessment process because it leads to improved learning.

Assessment criteria for the rubrics must be aligned with the outcomes to be achieved. Cangelosi (2000) suggests that three different domains must be assessed during group work: the cognitive, affective, and psycho-motor domains. The cognitive domain covers acquired knowledge and its processing (Vermeulen, 2002) and the ability of students to understand, analyse, apply, and evaluate the knowledge (Van der Horst & McDonald, 1997). When assessing the cognitive domain, Airasian (2005) believes that the facilitator needs to use assessment criteria that will assess the different levels of Bloom's taxonomy regarding students' low and high order thinking skills. The affective domain deals with students' beliefs, attitudes, and values (Vermeulen, 2002) and is normally assessed by members of the group and the facilitator during a group work activity. The members and facilitator observe ways in which students react to one another and the responsibilities they take on. The psycho-motor domain covers students' psychological and motor skills (Vermeulen, 2002). For assessing achievement in the psycho-motor domain, the physical activities/contributions of group members could be assessed. Facilitators often neglect these two domains because they are difficult to assess (Woolfolk, 1998; Vermeulen, 2002).

In order to assess groups adequately, Johnson and Johnson (2006) suggest that techniques fostering positive interdependence and individual accountability should be incorporated in the group work strategy. Positive interdependence is established when members of the group are so dependent on one another that one cannot succeed without the other. This is effected by coordinating one's efforts with the efforts of others to complete the task. Individual accountability is created when positive attitudes are in place that increase group members' feelings of responsibility and accountability. This can be done by assessing individual and group performance against predetermined assessment criteria, by communicating the results to individuals and the group, and by holding the members responsible for the quality of their individual contributions (Johnson & Johnson, 2006). Lejk and Wyvill (2001) also mentioned that self- and peer assessment are significant ways of ascertaining individual accountability and individual contributions.

## *Assessment Strategies Used in this Study*

In this section the specific assessment strategies used in this research will be elaborated on.

At the beginning of the course the facilitator explained the concept of pair programming and ways of assessment in detail to the students. Pair programming assignments were executed in a closed computer laboratory environment under supervision of the facilitator. Before each pair assignment, students were randomly allocated to pairs to ensure that they collaborated with a different programming partner every time. During the course, students had to complete four programming assignments in pairs. Ninety minutes were allocated to each programming session. Every twenty minutes, students rotated programming roles and ten minutes were reserved for assessment.

After the completion of every pair programming assignment, self-, peer- and facilitator-assessments were conducted by means of a rubric (see Table 1) containing specific assessment criteria. These rubrics were provided to students at the beginning of every pair programming assignment. Thus, assessment could take place formatively by the students and facilitator throughout the assignment.

**Table 1: Self-, peer- and facilitator-assessment rubric**

| ASSESSMENT CRITERIA | Not achieved | Achieved to a lesser extent | Achieved to a great extent | Totally achieved |
|---|---|---|---|---|
| **Cognitive outcomes*** | | | | |
| *Programme specific outcomes that changed from one week to the next, in accordance with: | | | | |
| Knowledge to be attained | 1 | 2 | 3 | 4 |
| Comprehension of knowledge | 1 | 2 | 3 | 4 |
| Application of knowledge | 1 | 2 | 3 | 4 |
| Analysis of facts and information | 1 | 2 | 3 | 4 |
| Synthesis of information in order to plan the programme | 1 | 2 | 3 | 4 |
| Evaluation of programme effectiveness | 1 | 2 | 3 | 4 |
| **Affective outcomes** | | | | |
| Members' contributions to collection of data and information | 1 | 2 | 3 | 4 |
| Interaction between partners | 1 | 2 | 3 | 4 |
| Responsibility of driver | 1 | 2 | 3 | 4 |
| Responsibility of navigator | 1 | 2 | 3 | 4 |
| Co-operation between members | 1 | 2 | 3 | 4 |
| Decision-making skills of members | 1 | 2 | 3 | 4 |
| **Psycho-motor skills** | | | | |
| Keyboard and mouse skills | 1 | 2 | 3 | 4 |
| Use of shortcut keys, textbooks and other resources | 1 | 2 | 3 | 4 |

The cognitive outcomes of each programming assignment differed from the others, in accordance with the cognitive outcomes specified in the study guide. All levels of Bloom's taxonomy were assessed by the cognitive assessment criteria. The affective assessment criteria assessed the interaction and collaboration between members, their individual responsibility, and their decision-making skills. These affective assessment criteria aimed at the fostering of a positive interdependence between the two members of the pair. Psycho-motor assessment criteria assessed keyboard and mouse skills as well as the use of shortcut keys, textbooks and other resources. Although Vermeulen (2002) has shown that the psycho-motor domain is hard to assess, in this study we have succeeded in its assessment due to the small number of students enrolled in the module. We were able to walk between the ten different groups and observe each group closely to assess their keyboard and mouse skills, shortcut keys, and the use of other resources. The facilitator gave formative feedback to the pair members before every new pair programming assignment, and, on its completion, individual tests were written. Only the cognitive outcomes required for the successful completion of the pair assignment were assessed by the individual tests. The aim of the individual tests was to determine whether both members of the pair had achieved the required cognitive outcomes and whether they were able to complete a similar programming assignment individually. The facilitator conducted these summative assessments. The individual tests as well as the self- and peer-assessments fostered individual accountability.

At the end of the course students were subjected to an individual and a pair practical examination. Both these examinations were conducted by the facilitator and covered the cognitive outcomes specified for the course.

In the empirical study that follows, we determine whether the specific assessment method for pair programming contributed to a more reliable way of assessing programming skills in pairs.

# Method

A non-experimental correlational design was used for this study. The research departed from a positivist perspective where the outcomes of the effect of the assessment strategies were based on the results of the quantitative data analysis. The aim of the research was to determine whether these specific assessment strategies could contribute towards a more reliable way of assessing individual abilities in pair programming situations. In order to determine quantitatively whether this aim has been achieved, the participants' individual marks were correlated with their pair assessment marks.

## *Participants*

All second-year students in teacher training (n=20) majoring in Information Technology (IT) in a Faculty of Education Sciences of a South African university took part in the research during 2006. The population consisted of 11 female and 9 male students, all in the age group 20 to 21 with the same cultural background and language proficiency.

## *Procedure*

For each pair programming assignment, the scores of the self-, peer- and facilitator-assessment per individual participant were summed and converted to a "Pair mark" out of 10 (See Table 2 for an example). The individual assessments were also converted to an "individual mark" out of 10. "Pair marks" and "individual marks" were then correlated.

**Table 2: Example: converting self-, peer, and facilitator assessment to a pair mark**

| | Self | | | Peer | | | Facilitator | | | Total | Converted |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cognitive | Affective | Psycho-motor | Cognitive | Affective | Psycho-motor | Cognitive | Affective | Psycho-motor | | |
| Student 1 | 20/24 | 22/24 | 6/8 | 18/24 | 20/24 | 8/8 | 20/24 | 22/24 | 8/8 | 148/168 | 8.81/10 |

Participants' scores obtained in the individual and pair practical examinations were also compared. These two examinations were not identical but of equal academic standard. The same learning outcomes were formulated for both of these examinations, and the assessment was conducted by the same examiner and moderated by the same moderator to ensure equal academic standards in both of these examinations. The content validity of the assessment rubrics, individual tests, and practical examinations were determined by experts in the field of computer programming who commented on the contents of the assessments.

### *Data Analysis and Statistical Techniques*

After each pair assignment and individual test, as well as the individual and pair practical examinations at the end of the course, the Spearman Rank Order Correlation technique was used to calculate the correlation coefficient between (1) average marks obtained in the pair and individual tests and (2) the average practical examination mark obtained in pairs and the average individual practical examination mark.

A coefficient of +1 indicates that the scores are perfectly positively correlated; a coefficient of -1 indicates a perfectly negative correlation; a coefficient of 0 indicates no linear relationship at all (Field, 2005). In this research the correlation coefficient was interpreted according to the following measure of effect size: 0.1 represents a small effect; 0.3 a medium effect; and 0.5 a large effect (Steyn, 1999).

## Results

Table 3 reflects the average marks of each assessment as well as the correlation coefficient and effect sizes obtained from the comparison between the average individual marks and the pair marks. Figure 1 shows that the average pair and individual assignment marks have both decreased from the first to the third assignment, but the average pair and individual assignment mark increased in the last assignment. Although participants performed better in the pair assignments in comparison with the individual assessment, the difference between the two assessments became smaller with each assessment.

In the comparison of the pair programming and the individual assessment marks, which was obtained after every pair assignment, it is interesting to see that after the first assessment no significant correlation existed between these two marks. However, after the second pair programming and individual assessment, a positive correlation (although of small effect size) was detected. After the third and fourth pair and individual assessments the correlation coefficients increased to values indicating medium effect.

**Table 3: Correlation between individual marks and pair mark**

| Assessment | Average individual marks | Average pair marks | Correlation coefficient (*r*) | Effect size |
|---|---|---|---|---|
| 1 | 8.30 | 9.44 | -0.14 | None |
| 2 | 7.7 | 9.33 | 0.18 | Small |
| 3 | 7.2 | 8.69 | 0.30 | Medium |
| 4 | 9.47 | 9.51 | 0.44 | Medium |
| Examination | 74.95 | 73.33 | 0.7 | Large |

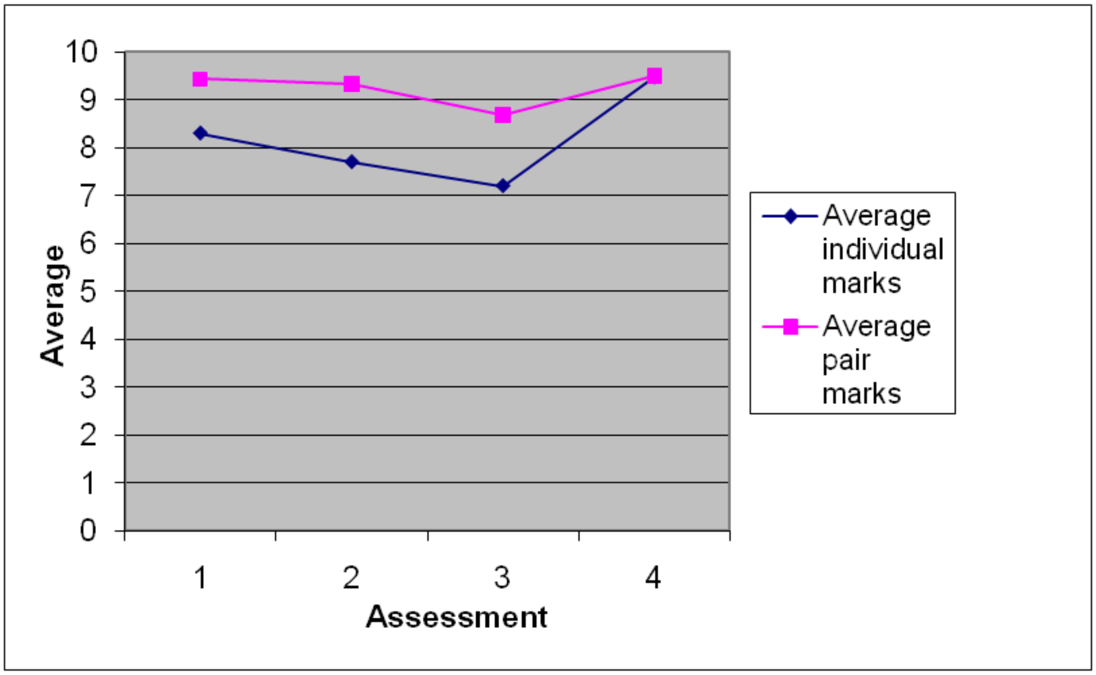


**Figure 1: Pair vs. individual average marks**

The second part of the investigation was completed at the end of the course to determine the correlation between the average practical examination mark obtained in pairs and the average individual examination mark. The average individual examination mark and the average pair mark were almost the same. A positive correlation coefficient of large effect, ($r = 0.7$) was obtained between the average marks of the practical pair examination and the practical individual examination.

# Discussion

The results of the first pair and individual assessments initially indicated that there was no correlation between the pair and individual marks, implying a difference between pair and individual marks. The averages of the pair assessment and individual test marks also indicated that the participants obtained higher marks in the pair programming assignment than in the individual test. This finding is consistent with findings from Daniel (2003), which indicate that students are initially unwilling to give poor marks to group members who did not contribute significantly towards the group task. However, as students became more familiar with this type of assessment, the correlation coefficients between the results of the pair and individual assessments increased, indicating that the marks for the individual and pair assessments became more aligned. As students became more familiar with this particular type of assessment strategy, they apparently became more realistic in assessing their own and their pair member's work as well.

The correlation between the individual and pair practical examination marks at the end of the course was anticipated in this research. Students became more used to this way of assessment and performed self- and peer-assessment more honestly and realistically. They could no longer solely depend on their fellow pair member to complete a pair programming assignment, because it would be reflected by the individual assessment marks which would indicate to the facilitator whether both members attained the outcomes specified for a particular pair assignment. The high correlation coefficient which was obtained between the practical individual and pair examinations indicates that the assessment strategies applied throughout the course resulted in a stronger commitment of both pair members to cooperate better when completing a programming task.

Morgan (2003) has indicated that students were not fond of group assessment because the students felt that their input was not assessed reliably. The findings in this study imply that the assessment strategies implemented in this research contributed positively towards a more reliable way of assessing individual abilities in pair programming situations. A more reliable way of assessment was done by assessing the input of the driver and navigator separately. Due to the assessment applied in this study the students have applied two general principals of group work (individual accountability and group processing) that, according to Johnson and Johnson (2006), are very important to take into account to perform any group work successfully. Individual accountability was applied by assessing each member of the pair individually. Because each student knew that his/her input would be assessed, each took up his/her responsibility as driver or navigator. Group processing took place when members of the pair communicated with each other about their progress in solving the problem and the extent to which goals were achieved. Peer assessment forced the members of the group to work together. In summary, because the different responsibilities were assessed apart from each other, individual accountability was fostered during pair work and effective cooperation between pairs took place. This may have contributed to the correlation between individual and pair assessment marks which were found from the third assignment onwards.

With the aid of the self-assessment rubric students had the opportunity of assessing themselves and, in the process, to monitor themselves and identify their own mistakes. The specific guidelines provided in the rubrics enabled the students to become the main agents of their own learning. Thus participants could monitor their own work and improve on their mistakes before the individual test and/or individual examination was written. The self-assessment rubrics could also contribute to the correlation found between the assessment marks that were obtained individually and the marks obtained in pairs.

It is important that the facilitators do not exclude themselves from the assessment process. After completion of the pair assignment the facilitator gave feedback to participants. The way the facilitator gave feedback was with the aid of the rubrics that also assigned a mark to their pair programming assignment. This type of feedback helped participants to improve their own learning and to guide them in future learning. Nonetheless, if the facilitator presents to participants a "good solution" and participants have the chance to compare their solution with that of the "good solution," they could improve their programming skills before writing the individual tests or individual examination. Feedback was then, as Earl (2003) states, part of the learning process. Therefore it helped with the correlation between individual and pair assessment.

# Conclusion

Although this study has provided some valuable insights, the results should not be unconditionally generalised due to the small number of students who participated in this study. It is recommended that the study should be replicated involving a larger sample of participants studies over a longer time period. More biographical information could also be useful for analysis of differences between male and female, different personality types, and age groups.

Nonetheless, this preliminary research has already shown that there could be a significant gain in assessing, not only the outcome of the assignment, but instead using different rubrics for self-, peer- and facilitator-assessment contributing to a more reliable assessment strategy for pair programming. The contribution of each individual was assessed and reflects the individuals' effort and ability to a great extent. The role of peer assessment as well as the honest assessment of one's own achievement of outcomes is very valuable to the learning process. It encourages honest reflection, both of one's own work and also of the work of others and creates a positive, collaborative culture in the class and a feeling of belonging. This fosters positive interdependence, which is important within any cooperative or collaborative learning environment. As soon as students get used to peer and self-assessment strategies, they became more honest in grading themselves and their pair members and as a result their pair and individual grading seems to correlate well with each other.

One benefit from peer-, self- and facilitator-assessment is that it adds significantly to the amount of feedback a student receives and can consequently contribute to a high level of achievement. Overall, the researchers believe that the results of the study support the use of various assessment strategies to assess the contribution of individual members of the pair during the pair programming activities.

These results can be of particular relevance to facilitators who contemplate implementing pair programming as a teaching strategy in the training of computer programmers, but who have some doubts about the reliable assessment of individual programming abilities while doing so. This could also be applied in general to assessment of any outcomes achieved in groups. Together these results show that a rubric-based tool for assessment of group work strongly supports the grading process.

# Acknowledgement

This research is based on work financially supported by the National Research Foundation (NRF) in South Africa. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors; the NRF does not accept any liability in this regard.

# References

Airasian, P. W. (2005). *Classroom assessment: Concepts and applications* (5th ed.). New York: McGraw-Hill.

Adams, A., Lubega, J., Walmsley, S., & Williams, S. (2004). The effectiveness of assessment learning objects produced using pair programming. *Electronic Journal of E-learning, 2*(2), 247-256. Retrieved: August, 2005, from http://www.ejel.org/volume-2/vol2-issue2/v2-i2-art1-adams.pdf

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, pp. 139-148.

Cangelosi, J. S. (2000). *Assessment strategies for monitoring student learning.* Menlo Park: Longman.

Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in Higher Education, 5*(2), 243-255.

Daniel, C. (2003). Experience with pair programming at a small college. *Journal of Computing Sciences in Colleges, 19*(1), 20-29.

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, California: Corwin Press.

Ellis, G. (2001). Looking at ourselves – Self-assessment and peer assessment: Practice examples from New Zealand. *Reflective Practice, 2*(3), 298-302.

Field, A. (2005). *Discovering statistics using SPSS.* (2nd ed.). London: SAGE.

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development, 20*(1), 53-70.

Johnson, D. W., & Johnson, F.P. (2006). *Joining together: Group theory and group skills* (9th ed.). Boston: Pearson.

Knight, J. (2004). Comparison of student perception and performance in individual and group assessments in practical classes. *Journal of Geography in Higher Education, 28*(1), 63-81.

Lambert, D., & Lines, D. (2000). *Understanding assessment: Purposes, perceptions, practice*. London: Routledge Falmer.

Lejk, M., & Wyvill, M. (2001). The effects of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment & Evaluation in Higher Education, 26*(6), 551-561.

Liebovich, B. J. (2000). Children's self-assessment. *Proceedings of the Lilian Katz Symposium.* Issues in early childhood education: Curriculum, teacher education, and dissemination of information (pp. 237-244). Champaign.

Luckett, K., & Sutherland, L. (2000). Assessment practices that improve teaching and learning. In S. Makoni (Ed.), *Improving teaching and learning in higher ducation: A handbook for Southern Africa* (pp. 98-130). Johannesburg: Witwatersrand University Press & The Higher Education Research and Development Society of Australasia.

Marneweck, L., & Rouhani, S. (2002). Continuous assessment. In: M. Jacobs, N. Gawe, & N. Vakalisa (Eds.), *Teaching-learning dynamics: A participative approach for OBE* (2nd ed., pp. 278-327). Johannesburg: Heinemann

McDowell, C., Hanks, B., & Werner, L. (2003). Experimenting with pair programming in the classroom. *ITiCSE,* pp. 60-64.

Morgan, B. M. (2003). Cooperative learning in higher education: Undergraduate student reflections on group examinations of group grades. *College Student Journal, 37(1),* 1-9.

Naude E., & Hörne, T. (2006). Cheating or 'collabarative work': Does it pay? *Issues in Informing Science and Information Technology, 3,* 459-466. Retrieved from http://informingscience.org/proceedings/InSITE2006/IISITNaud223.pdf

Parsons, D. (2004). Justice in the classroom: Peer assessment of contributions in group projects. *Proceedings of the 17th NACCQ, Christchurch,* 145-152. Retrieved April 22, 2008 from http://www.naccq.ac.nz/conference05/proceedings_04/parsons.pdf

Preston, D. (2005). Pair programming as a model of collaborative learning: A review of the research. *Journal of Computer Science in Colleges, 20*(4), 39-45.

Steyn, H. S. (1999). *Praktiese beduidendheid: Die gebruik van effekgroottes* [Practical significance: The use of effect sizes]. Potchefstroom: Potchefstroomse Universiteit vir Christelike Hoër Onderwys.

Tomayko, J. E. (2002). A comparison of pair programming to inspection for software defect reduction. *Computer Science Education, 12*(3), 213-222.

VanDeGrift, T. (2004). Coupling pair programming and writing: Learning about students' perceptions and processes. *Special Interest Group on Computer Science Education,* pp. 2-6.

Van Der Horst, H., & McDonald, R. (1997). *Outcomes-based education: A teacher's manual*. Pretoria: Kagiso.

Vermeulen, L. M. (2002*). Kurrikulum 2005: Skool- en klaskamerorganisasie en assessering vir die senior fase (Graad 7-9)* [Curriculum 2005: School and classroom organisation and the assessment of the senior phase (Grade 7-9)]. Vanderbijlpark: Van Schaik.

Verhaart, M., Hagen, K., & Giles, O. (2005). Best practice assessment methods for evaluating an individual's performance in group work. *Proceedings of the 18th Annual Conference of the National Advisory Committee on Computing Qualifications, New Zealand,* 113-120. Retrieved 22 April, 2008 from http://www.naccq.ac.nz/conference05/proceedings_05/papers/verhaart.pdf

Williams, L., & Kessler, R. (2001). Experiments with industry's "pair-programming" model in the computer science classroom. *Computer Science in Education, 11*(1), 7-20.

Williams, L., & Kessler, R. (2003). *Pair programming illuminated*. Boston: Pearson.

Williams, L., & Upchurch, R. L. (2001). In support of student pair-programming. *Special Interest Group on Computer Science Education*, *2*(1), 327-331.

Williams, L., Wiebe, E., Yang, K., Ferzli, M., & Miller, C. (2002). In support of pair programming in the introductory computer science course. *Computer Science Education, 12*(3), 197-212.

Woolfolk, D. (1998). *Teaching for learning*. Retrieved January 24, 2006, from: http://165.193.130.33/textlink/In/RE/ Woolfolk/ebook/Wool.d.html

# Biographies

**Jan Hendrik Hahn** is currently a lecturer in Computer Science Education at the North-West University. He is interested in the research area of cooperative learning. Previously he taught Computer Science at a High School for grades 10 to 12.

**Prof Elsa Mentz** is a professor in Computer Science Education in the Faculty of Education Sciences at the North-West University (Potchefstroom Campus). In her research she focuses on teaching and learning of computer programming skills. She holds a National Research foundation grant on group work in Computer Science Education. She has been actively involved in the training of IT teachers for the past twelve years.

**Prof. Lukas Meyer** is attached to the School of Education at the North-West University, Potchefstroom Campus. He is a registered educational psychologist and he teaches modules in educational psychology to pre-graduate and graduate students. His academic interests revolve around psychological and educational assessment and intervention. He has published a number of articles in national and international journals and presented various papers at national and international conferences on psychological and educational assessment. At the moment he is the programme leader of the MEd and PhD programmes in the Faculty of Education Sciences at the Potchefstroom Campus of the North-West University.